SESAR Engage KTN – catalyst fund project final technical report

Project title:	Data-Driven Trajectory Imitation with Reinforcement Learning
Coordinator:	University of Piraeus Research Center
Consortium partners:	Boeing Research and Technology Europe
Thematic challenge:	TC2 Data-driven trajectory prediction
Edition date:	23 November 2020
Edition:	1.1
Dissemination level:	Public
Authors:	George Vouros / University of Piraeus Research Center

This project has received funding from the SESAR Joint Undertaking under the European Union's Horizon 2020 research and innovation programme under grant agreement No 783287.

1. Abstract and executive summary

1.1 Abstract

The objective of this project was to present algorithms for data-driven imitation of trajectories, following deep reinforcement learning techniques towards enhancing our trajectory prediction abilities.

We aimed at building a data-driven approach in which the learning process is (a) an imitation process, where the algorithm tries to imitate "expert", demonstrated trajectories, (b) exploiting raw trajectory data, enriched with contextual data (e.g. weather conditions etc) and (c) based on reward models (for producing trajectories in high-fidelity) that are learned during imitation.

In this project we have focused on predicting trajectories at the pre-tactical phase of operations, considering mainly the interests of airspace users, i.e. without considering, for instance, air traffic controllers' measures, availability of routes or sectors congestions.

This document presents the methodology followed during the project, outcomes, findings and contributions made, as well as lessons learnt and future activities.

The major contributions are as follows:

(a) We devised a general framework for the prediction of trajectories in which deep imitation and reinforcement learning methods play a major role, together with methods selecting important features for decision making and future trajectory classification methods;

(b) We developed and evaluated state of the art deep imitation learning techniques for predicting trajectories in the aviation domain, showing their potential for highly accurate prediction results, especially in long trajectories with multiple patterns / modalities, and in cases where the demonstrated trajectories are few.

1.2 Executive summary

Building on the knowledge gained from the DART SESAR ER-2 project on enhancing our trajectory prediction abilities, and aiming at building a straightforward data-driven approach following deep reinforcement learning techniques, we approached the learning process (a) as an imitation process, where the algorithm tries to imitate 'expert', demonstrated trajectories, (b) exploiting raw trajectory data enriched with contextual data that provide information relevant to the evolution of trajectories, and (c) based on reward models that are learned during imitation.

It is the objective of this project to learn models for the evolution of trajectories, exploiting historical, demonstrated trajectories, which (models) can be used for predicting future trajectories. Towards this goal we formulated the trajectory imitation problem as a Markov Decision Process and applied deep reinforcement learning (DRL) methods.

Contributions made are as follows:

a) We approached the flight trajectory prediction problem as an imitation problem, using DRL models learnt from historical data: According to our knowledge, this is the first time that these state-of-the-art machine learning techniques are used for the prediction of the trajectories. We delivered two methods, which have been evaluated using short (single-FIR) and long (multiple-FIR) trajectories, with very promising results.

b) We have used advanced algorithms for identifying the features relevant to airspace users for evolving flight trajectories, towards learning their reward model. Tactical interactions and conflicts between trajectories are not explicitly addressed.

c) We have built a generic methodology and computational framework for the prediction of trajectories, comprising methods for identifying patterns of demonstrated trajectories (modalities of behaviour), identifying the features relevant to following different modes of behaviour, classifying future trajectories, and predicting trajectories using DRL methods.

Contributions to the ATM Master Plan are as follows:

a) Increased abilities for flight prediction and planning, by means of learning models of trajectories planned and flown by airspace users.

b) Improved operations productivity via contributions to improved flight prediction and planning.

We believe – based on the outcomes produced – that this project will serve as a catalyst for the use of deep reinforcement learning methods to predict short/long trajectories in ATM, either at the pre-tactical or at the tactical stages of operations, following an imitation learning approach: The project surely matured the ideas and advanced the data-driven trajectory prediction methods produced in DART, as well as improved the state of the art in data-driven trajectory prediction.

2. Overview of catalyst project

2.1 Operational/technical context

The current Air Traffic Management (ATM) system worldwide has reached its limits in terms of predictability, efficiency and cost effectiveness. Nowadays, the ATM is based on an airspace management paradigm that leads to demand imbalances that cannot be dynamically adjusted.

With the aim of overcoming the ATM system deficiencies, different initiatives, dominated by SESAR[1] in Europe and Next Gen[2] in the US, have promoted the transformation of the current environment towards a new trajectory-based ATM paradigm. This paradigm-shift changes the old-fashioned airspace management to the advanced concept of Trajectory Based Operations (TBO). In the future ATM system, the trajectory becomes the cornerstone upon which all the ATM capabilities will rely on. The trajectory lifecycle describes the different stages from the trajectory planning, negotiation and agreement, to the trajectory execution, amendment and modification.

The proposed transformation requires advanced aircraft trajectory prediction capabilities, supporting the trajectory lifecycle at all stages efficiently. Making high-fidelity plans of trajectories to be flown in an early phase of operations, should allow predicting ATM network status effectively, reduce factors of uncertainty and boost the effectiveness of operations' planning. In addition, advances towards this direction should support effective decision making and optimization of resources' exploitation during operations time.

This project, aims at data-driven trajectory prediction, building on results and experience from state-of-the-art data-driven trajectory prediction methods, and considering that Reinforcement Learning techniques inherently deal with trajectories, formed as policies in an action-state space. It particularly focuses on using imitation learning methods exploiting deep reinforcement learning techniques to predict trajectories at the pre-tactical phase of operations, considering mainly the interests of airspace users, i.e. without considering, for instance, air traffic controllers' measures, availability of routes or sectors congestions.

Considering the Engage Thematic Challenges, the proposed research contributes towards increasing trajectory predictability, as it aims to imitating flown trajectories by inferring high-fidelity trajectory models, incorporating various features affecting flights. Increasing predictability via highly accurate and highly potential-to-happen trajectories devised at the planning stage, reduces buffers and uncertainty in operations.

In addition to these, this research contributes – in an indirect way- towards collaborative decision making, supporting better planning of operations from Airspace Users, without explicitly considering conflicts among trajectories.

---

1 SESAR 2020, http://www.sesarju.eu

2 NextGen, https://www.faa.gov/nextgen

2.2 Project scope and objectives

The goal of this project is to learn high-accurate models for predicting trajectories with low Root Mean Square Error (RMSE) in 4D along the produced trajectory, in comparison to the actual (flown) trajectory.

We aim to develop and evaluate advanced deep reinforcement learning methods that are trained to imitate trajectories, treating these trajectories as rollouts of expert policies performed in an action-state space. By doing so, we followed a supervised leaning approach, and treat historical data as data provided by "experts" that a machine learning algorithm should exploit to learn the corresponding policies comprising actions for transiting between positions in the 3D space through time. Then, learned models can be exploited to predict trajectories.

"Experts" in our case can be all those entities that affect the trajectory: I.e. the basic assumption behind this research is that the flown trajectory codifies somehow the decisions of different stakeholders and thus, a data-driven algorithm will learn a model that incorporates these features, imitating experts to producing trajectories, given sufficient data to do so. This assumption has been verified by data-driven methods towards trajectory prediction developed and evaluated in the DART project.

In this project we have focused on the prediction of trajectories at the pre-tactical phase of operations, considering mainly the interests of airspace users, i.e. without considering, for instance, air traffic controllers' measures, availability of routes or sectors congestions. Thus, we focused mainly on evaluating the proposed methods in long (multiple-FIR) trajectories, where route charges and weather conditions play a major role, also considering short (single-FIR) trajectories but without considering any of the constraints and measures that apply during the tactical phase. It must be noted, that indicative results show that the devised methods can also be used at the tactical stage of operations, although further exploration is needed towards this.

While imitation is advantageous when the trajectories are constructed by experts, we may be tempted to generalize beyond the cases considered during learning (which is completely justified and highly-desired) and learn policies that are substantially better from those produced by experts: In our case, we need to learn high-fidelity models based solely on experts' policies, i.e. from real trajectories flown. This is crucial here, given that we need to imitate experts without going further on optimizing during the learning process. Based on that, exploiting the high-fidelity experts' models learned, we will be able to produce trajectories that are close to those produced in reality.

Reinforcement learning techniques inherently deal with trajectories, which are produced by policies in an action-state space. Such methods have been used in predicting aircraft trajectories in various domains, but their use in the aviation domain has only started in the DART project.

Therefore, building on experience and knowledge gained in DART, we aim at building a straightforward but novel approach in which the learning process is (a) an imitation process, where the algorithm tries to imitate demonstrated trajectories, (b) exploiting raw trajectory data enriched with contextual features, and it is (c) based on reward models that are learned during imitation.

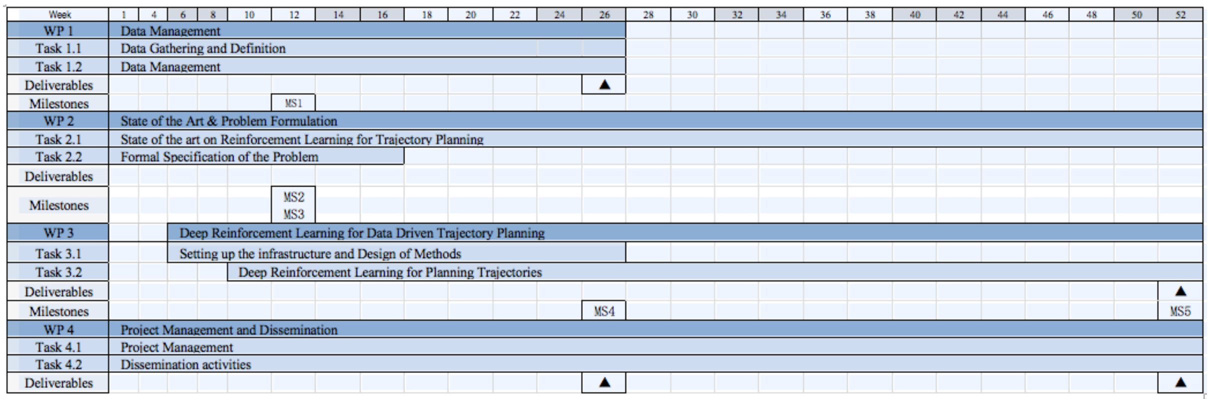
Exploiting raw trajectory data is advantageous considering that, (a) methods learn directly policies on producing trajectories rather than producing commands generating trajectories as in the case of exploiting aircraft intent, (b) we do not need to incorporate any model-based prediction method into the process (at it would be for instance when exploiting aircraft intent).

Based on the problem formulations provided during the project, our main aim was to develop methods learning in a continuous action-space, without discretizing actions or states' features. In doing so, staying in the deterministic case and avoiding learning a state-transition model, allows us to develop more stable and efficient learning methods. In addition to this, we explored advanced methods in discrete action spaces and compared these to methods exploiting continuous actions.

An additional but crucial task in this project was to develop a framework and methodology providing the overall infrastructure and architecture for predicting trajectories, enabling us to performing training and testing tasks with the machine learning algorithms used, towards increasing our trajectory prediction abilities.

2.3 Research carried out

The work plan comprised four work-packages, the fourth being the project management and dissemination workpackage, and are as follows:



Milestone number	Milestone name	Due Date (week/month)[12]	Means of verification
MS1	Data Sets Specified and Gathered	W12/M03	Data sets have been gathered and described in a report and they have been made accessible to the participating entities
MS2	State of the Art Methods Reviewed	W12/M03	State of the art methods have been studied and the most promising method(s) have been identified. Updates will follow until nearly the completion of the project.
MS3	Formal Problem Specification	W12/M03	The problem specification is documented and updates (considering feedback from preliminary evaluation results) will follow up to W16 (M04).
MS4	1st version of deep Reinforcement Learning methods	W26/M06	Deep Reinforcement Learning Methods for Trajectory Planning with initial evaluation results reported.
MS5	Final version of Deep Reinforcement Learning Methods	W52/M12	Deep Reinforcement Learning Methods for Trajectory Planning with final evaluation results reported.

Below we summarize the work being done per workpackage (WP), the methodology applied, tools and methods used / developed, and outcomes produced.

WP1 “Data Management”

The objectives of this WP are as follows:

1. Gather data sets required

2. Process and associate data from different data sets

3. Manage and curate data sets

During the project, we have gathered data concerning (a) short trajectories executed in the Spanish airspace from three months in 2019: January, April and July; and (b) long trajectories crossing multiple FIRs in the European continent during January and July of 2019. The periods (months) have been selected deliberately to reflect different traffic and contextual (e.g. weather) conditions.

Together with surveillance data we have gathered data regarding weather conditions (actual and forecasted), sector configuration data, flight plan data regarding flown trajectories, and these were also combined with data on airports, METAR and TAF data. Of course, all these datasets are aligned in space and time with surveillance data [1].

These datasets were provided by Boeing Research and Technology Europe, who gave UPRC access to their data stores via their ADAPT system.

In addition to these, we have also complemented the data sets with data on MTOWs (given the existing aircraft types), as well as en-route charge rates for European FIRs.

Specifically, gathered datasets consist of:

Weather data: These comprise NOAA data, as well as ASOS(METAR) and TAF data. NOAA data concern the three months (January, April and July 2019) inside the Spanish airspace i.e. the Spanish part of the Iberian Peninsula plus the Canary Islands. In addition, we have collected NOAA data for two months (January and July 2019), for the relevant areas for the trajectories from London to Rome and from Helsinki to Lisbon. ASOS data has been gathered for 47 Spanish aerodromes for January, April and July 2019, as well as LHR, FCO, HEL and LIS for January and July of the same year. Finally, TAF data is available for 44 Spanish airports during the months selected, as well as the four previously mentioned airports used for the multiple FIRs trajectories.

Surveillance (Radar) data: This is data provided by Flightradar24. It comprises aircraft positional messages which share some fields with MSG ADS-B. This data is available for January, April and July of 2019, covering the entire Spanish airspace. Moreover, Flightradar24 data is available for flights from London to Rome and from Helsinki to Lisbon for January and July of 2019.

Sector Configuration data: This is AIXM data covering 2019 with a small number of days missing.

Flight Plan data: DDR ALLFT+ is the ATM type of dataset covering flight plan information from EUROCONTROL Demand Data Repository (DDR). January 2019 is available for the entire Spanish airspace.

These datasets are described in D1.1. "Datasets Description" [1], together with a description of links (associations) between them.

Regarding Task 1.2, we used tools that we have developed in other projects (DART[3] and datAcron[4]) to parse datasets and developed methods for parsing new datasets, as well as for (a) associating subsets of surveillance data with other datasets, producing enriched trajectories, and (b) clustering enriched trajectories. Specifically, we enriched surveillance data (i.e. flown trajectories) between specific origin destination pairs with weather (NOAA, METAR), Sector Configuration data (for the short trajectories, only), as well as with costs regarding en-route charges per FIR (for the long, multiple FIR trajectories) according also to [8], as well as with aircraft and airlines attributes.

[3] http://dart-research.eu/

[4] http://datacron-project.eu/

WP2 “State of the Art Review & Problem Formulation”

The objectives of this WP are as follows:

1. Study thoroughly state of the art techniques on (data-driven) Reinforcement Learning for trajectory imitation and planning,

2. Specify the trajectory imitation and planning problem formally, as an MDP.

State of the art techniques reviewed and compared, so as to provide evidence on the novelty and significance of the developments in this project.

Specifically, we have completed the State of the Art deliverable D2.1 which is available at [2], reviewing prominent data-driven trajectory prediction approaches and important state of the art reinforcement learning techniques closely related to imitating trajectories.

Also, regarding task T2.2 we formulated the problem as an MDP to be solved by reinforcement learning methods. We actually provided two alternative formulations reported in the Deliverable 2.1 [2], the one being suitable for trajectory predictions in continuous state-action spaces, while the other discretizing the actions available to the trajectory predictor. Both formulations have been tested using the corresponding methods developed in WP3.

WP3 “Deep Reinforcement Learning for Data Driven Trajectory Planning”.

This WP comprises the main activities of this project, aiming to develop deep Reinforcement Learning methods towards imitating and predicting flight trajectories.

Specifically, the concrete objectives of this WP are as follows:

- Setting up the infrastructure necessary for experimenting with deep Reinforcement Learning methods, and provide a design of the overall framework and individual method;
- Develop and evaluate a deep Reinforcement Learning method for learning trajectories by imitating demonstrated trajectories, using data provided by WP1.

Specifically, regarding task T3.1 we have provided a general methodology and computational framework for predicting trajectories:

A high-level description of the pipeline of tools and methods used is provided in Figure 1. The first stage comprises raw, demonstrated trajectory enrichment and pre-processing, the second stage comprises demonstrated trajectory clustering, while the third stage includes feature selection and future trajectory classification, and finally, trajectory prediction via deep imitation learning techniques.

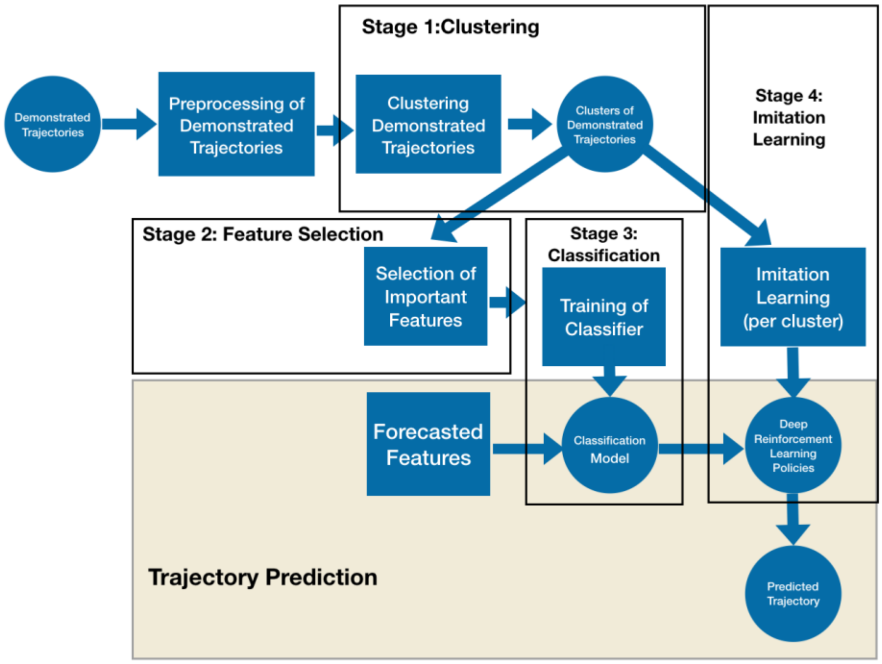


Figure 1 – The stages of the methodology proposed, realized by the infrastructure devised during the project. All stages from 1 to 4, as well as the pre-processing of the historical demonstrated trajectories are offline stages, providing classification models and policies learnt. These models and policies are exploited by the trajectory prediction stage for online trajectory prediction.

The pre-processing of the surveillance data and its association with other data sources is an offline stage and includes two main stages. First, we interpolate points within trajectories to ensure constant temporal intervals between all trajectory points. Data cleaning is the second stage. Here we analyze the trajectories and drop out outliers which would result in unstable behavior during the learning process. There are two distinct categories of trajectories dropped. The first one includes incomplete trajectories that may start or finish away from the origin or the destination airports. The second one includes flights that show inconsistent behavior, for example covering a significant distance within an unreasonable amount of time, resulting in velocity much greater than the maximum speed of the corresponding aircraft model.

What we need to do towards automating the data-driven trajectory prediction process is to detect distinct patterns of trajectories, identifying also the features that distinguish these patterns. Then, we can by imitation learn a distinct policy per class of trajectories, i.e. for those trajectories following a specific pattern of behavior. This can make the learning process much more efficient and effective in contrast to training a single model, considering all possible trajectories with all different modalities. However, to predict a single trajectory we need to know which policy to apply, thus, the mode of behavior it will most probably follow during the period it will be executed. One solution to this is to identify and forecast the contextual features that may impact the mode that a trajectory will follow. This (future trajectory) classification step is thus restricted to those features, which they do distinguish between different modes of behavior, and which can be forecast or can be known the pre-tactical stage of operations.

Thus, the trajectory prediction approach that we propose incorporates a trajectories clustering step (identifying different modes of behaviour), a future trajectory classification step (selecting the most probable mode of behaviour), and finally a trajectory imitation step (predicting the trajectory evolution given a specific mode of behaviour).

As specified in Figure 1, (historical) trajectories clustering is an offline process that it uses the full set of trajectories' enriching features. The main difficulty of this task is that the appropriate number *K* of clusters is unknown. The problem of determining K can be transferred to a silhouette coefficient maximization problem. The computation of the silhouette coefficient needs only pairwise distances and the calculation of clusters' centroids is avoided. We have used two alternative approaches for clustering data trajectories exploiting distance measurements (normalized Mean Squared Error and normalized Dynamic Time Warping): a modification of the k-means algorithm that constitutes a partitioning clustering procedure, and the agglomerative hierarchical clustering scheme that provides a bottom-up structure of a dataset.

Then, feature selection aims to identify the features that are more relevant to deciding on the mode of behavior to be followed. Feature selection reduces the dimensionality and aims to support (future) trajectory classification in a reduced space, in high prediction accuracy. Important features may be different from those used during imitation learning and should be forecasted / decided at the pre-tactical stage of operations.

Specifically, the feature selection problem is as follows:
Given,

- a set F of features,
- a set of K clusters (classes representing different modes of behaviour) $C_{E,l}$, l = 1...K, of trajectories enriched with these features,
- a set of instances (x_i, y_i) where $x_i \in R^d$ denotes a feature vector and $y_i \in \{1,\ldots,K\}$ is the corresponding class label,

we aim to determine the set of most important features that should be included in the set T^f of contextual features of trajectories (let that be T^{f*}), and the set of important features of enriched states at specific "landmark" positions S^f that the future trajectory will cross (let this new set of enriched states be S^{f*}), so as to increase the efficiency and accuracy of classifying a future trajectory to one of the K classes.
As the set of important features may change between origin-destination pairs we need a method that automates the feature selection process, raking the features in F fed into the processes.
Towards this goal we selected two state of the art methods:

(a) The Neighbourhood Component Analysis (NCA) [21] method aiming to reduce the dimensionality of a classification problem by learning a distance measure to be used in the KNN classification algorithm, and
(b) the LIME method [0] aiming to explain the predictions of any classifier by learning an interpretable model around a prediction and providing explanations in a non-redundant way, ranking the features used for classification based on the their importance.

Clustering and feature selection steps are performed prior to the imitation learning training stage in an offline way and can be automated. Imitation learning algorithms are trained using clustered data, so as to learn a single policy per cluster. The classification model and the policies learnt via data-driven trajectory imitation algorithms, are used to provide online trajectory predictions.

Among the important features we have identified meteorological parameters, at specific points in the trajectories, as well contextual features that concern the cost, duration, distance flown etc. of the trajectory. However, meteorological features are shown to be important in seasons where weather conditions may be severe (in January), while they do play a less important role in good weather conditions or in seasons whether other factors prevail (e.g. traffic). However, our analysis does not show a decisive, consistent role that these features play for selecting the mode to be followed in multi-FIR-trajectories.

Thus, in all cases, the most important common features are those concerning the cost of the trajectory, as a function of the en-route charges of the FIRs crossed, together with the distance flown, also compared to the minimum distance between origin-destination airports.

In any case, either by forecasting the important features (which results on a cyclic prediction process between predicting trajectories and costs), or by making specific choices for the range of their potential values, classification accuracy of the future trajectories based on these features is very high: Thus, our classifiers can predict the mode of behaviour to be followed, and inform the reinforcement learning model to select the proper policy model for predicting the trajectory evolution.

Details on the individual methods and the necessary computational infrastructure implementing the framework shown in Figure 1 has been set up during the project: This is the outcome of T3.1 and is documented in D3.1 [3].

Regarding task T3.2, based on the data-driven deep reinforcement learning state of the art methods identified, as well as on the problem formulation, we have built two methods exploiting raw trajectory data in association to other data, as provided by WP1, for learning how to imitate trajectories.

The first method, based on GAIL [17], directly learns the optimal policy from expert demonstrations, quite efficiently, since it does not need to explicitly derive a reward function that will be used by a reinforcement learning method to derive a policy, nor it makes any assumptions regarding the form of the reward function, which is fitted using a discriminator neural network. Actually, we have used a variation of GAIL – especially for long trajectories- emphasizing on states visited by trajectories, rather than on actions performed by agents, very closely to the GAIfO [19] approach. This method uses an architecture similar to Generative Adversarial Networks to find the best policy that imitates the demonstrated trajectories.

AppLearn, the second method devised, is based on deep Q-learning methods, using an apprenticeship learning approach [6]: ApplLearn allows us to directly derive new policies according to the expert policy, as this is demonstrated by means of the historical trajectories. The problem is formulated as an MDP/R, i.e. MDP without knowing the “true” reward function, and the assumption is that the apprentice is trying to optimize this unknown reward function which can be expressed as a linear combination of its features. Then, using the learnt reward a Deep Q-Network is trained with the fully known MDP. The underlying assumption regarding the reward function is that this function is approached with a linear approximation of the enriched trajectory features.

As part of the work done in WP3, the overall trajectory prediction framework, as well as constituent methods, with emphasis on deep imitation learning approaches used – GAIL and AppLearn – have been thoroughly evaluated in the following cases:

- Single FIR (short) trajectories:
    - BCN-MAD (2016): This includes 528 trajectories flown from BCN to MAD during April 2016. The trajectories were clustered into 2 clusters with 250 and 278 trajectories, respectively.
- Multi FIR (long) trajectories:
    - LHR-FCO (2019):
        - January: This includes 218 trajectories flown from LHR to FCO during January 2019. The trajectories were clustered int 2 clusters with 23 and 195 trajectories, respectively.

- July: This includes 242 trajectories flown from LHR to FCO during July 2019. The trajectories were clustered int 3 clusters with 4, 19 and 219 trajectories, respectively.

  - HEL-LIS (2019):

    - January: This includes 49 trajectories flown from HEL to LISBON during January 2019. The trajectories were clustered int 3 clusters with 4, 3 and 42 trajectories, respectively.

    - July: This includes 55 trajectories flown from HEL to LISBON during July 2019. The trajectories were clustered int 3 clusters with 2, 9 and 44 trajectories, respectively.

Results, show the following:

First, the clustering methods provide a set of classes, identifying potential modes that future trajectories may follow. These, together with the feature selection and classification methods devised can predict the mode that a future trajectory will follow in high accuracy (greater than 95%), given features' forecasted values in few landmark spatio-temporal positions (e.g. weather in the destination airport, or en-route charges for a specific modality).

Regarding the imitation learning techniques developed and evaluated, results show the superiority of GAIL in all cases, given also that AppLearn is highly competitive, even for very long trajectories with very few demonstrated trajectories. Indeed, GAIL has achieved the best results among all methods (AppLearn and state of the art data-driven methods), even in cases with a small training set, also w.r.t. the average distance of the trajectories between each origin-destination pair considered. These results are detailed in D3.1 [3], in terms of RMSE and Cross/Along Track Errors, as well as Vertical Errors, errors in Estimated Time of Arrival and errors in total cost with respect to en-route charges of the FIRs crossed.

To show what has actually achieved in this project, and in comparison with state of the art trajectory prediction methods, we show in Figure 2 the proportion of mean RMSE increase compared to GAIL, as reported by all methods, w.r.t. the proportion of the average distance between the origin-destination airports, compared to the BCN-MAD pair :

- The DART method (indicated as BCN-MAD (DART) in Figure 2), although a constrained method that considers the flight plans during predictions, increases the mean RMSE by 1.13 for the same origin destination pair (i.e. no increase of the average trajectory distance).
- The HMM method proposed in [7] (indicated by the pair ATL-MIA in Figure 2) increases the mean RMSE by 1.02 for an increase of the average trajectory distance by 1.98: This shows a highly competent method, however there are no results from longer trajectories: A more thorough study between these methods is needed.
- On the other hand, the deep method proposed in [18] (indicated by the pair IAH-BOS in Figure 2), increases the mean RMSE by 7.27 for an increase of the average trajectory distance by 5.31. To better understand how this score is compared to what it is reported by GAIL, this is an increase to the mean RMSE by 0.87 for an increase of the average trajectory distance by 0.76, compared to the HEL-LIS (January) case, and an increase to the mean RMSE by 2.03 for an increase of the average trajectory distance by 1.78, compared to the LHR-FCO (January) case. However, the vertical error reported by [18] (i.e. 2800ft) is large compared to what is reported by GAIL.

It must be noted that authors are fully aware that the differences in predictions between GAIL, AppLearn and DART and methods proposed in [7] and [18] might be due to characteristics of the different airspaces/types of environment (particularly for US or European regions and intra- or inter-continental flights), which may impose difficulties to draw conclusions if the same routes are not analysed with the different methods. Thus results should be considered only as providing evidence to the potential of the methods investigated, rather than on proving superiority of our methods relative to others.

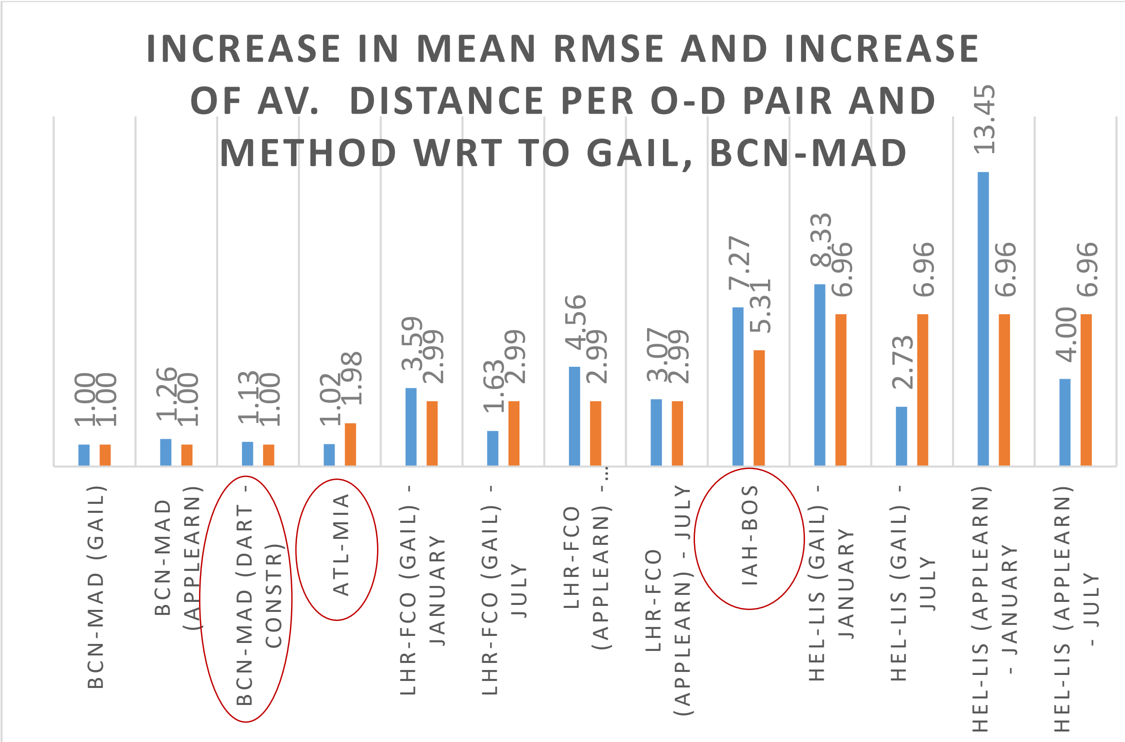


Figure 2 – Comparison between GAIL, AppLearn and state of the art methods in terms of mean RMSE (blue bars) and trajectories' average distance (orange bars) w.r.t to the BCN-MAD pair. The HMM method proposed in [7] is indicated by the pair ATL-MIA considered there, and the deep method proposed in [18] is indicated by the pair IAH-BOS.

WP4 "Project Management & Dissemination Activities"

Activities in this WP aim at monitoring and controlling of all tasks, coordination of interfaces between the activities, scheduled planning and status control of all activities, resource and finance control of the entire project, risk identification and mitigation control, dissemination of results, interaction with stakeholders, and liaison with Engage KTN.

Partners had 3 meetings and continuous discussions during the project, towards

- Clarifying data access and data exploitation issues;
- Understanding the potential and limitations reflected in initial results regarding deep reinforcement methods for trajectory imitation;
- Deciding on the pairs of origin-destination pairs that should be used for evaluating the devised framework and trajectory prediction methods and refining the scope and planning of the project.

Regarding the last issue, partners clarified the focus on long trajectories crossing multiple FIRs, also w.r.t. other operational features regarding en-route cost charges per European FIR. In doing so, we focused on increasing the trajectory prediction horizon, while at the pre-tactical stage of operations.

This necessitated the generalization of the prediction methodology devised, towards introducing a more general feature selection and future trajectory classification approach, in contrast to the simple approach used in short, single FIR trajectories.

Therefore, the scope and planning of the project focused on

- Generalizing and automating the methodology for trajectory prediction at the pre-tactical stage, focusing on the interests of airspace users, incorporating trajectory clustering methods, advanced feature selection, future trajectory classification, and imitation learning methods.

- Tuning and testing the deep imitation learning approaches to the prediction of the long trajectories.

2.4 Results

Contributions made are as follows:

We approached the flight trajectory prediction problem as an imitation problem, using DRL methods that learn models from historical data: According to our knowledge, this is the first time that these state-of-the-art machine learning techniques are used for the prediction of flight trajectories, proving their high potential to predicting trajectories crossing multiple FIRs, in long time horizons.

We delivered two imitation learning methods, challenging their potential and assumptions towards predicting short (single-FIR) and long (multiple-FIR) trajectories, thus evaluating their abilities to increase the prediction horizon at the pre-tactical stage of operations, w.r.t. other state of the art methods. In addition to these, our results show that the prediction methods can be also used during the tactical stage in an online way – although further research is needed towards this target.

Results show that the GAIL-based method prevails AppLearn and state of the art data-driven trajectory prediction methods, in all cases considered, even in cases with long trajectories and few training examples (i.e. demonstrated trajectories). This is revealed by the low mean RMSE reported by that method in comparison to the mean RMSE reported by other methods w.r.t the average distance of the predicted trajectory.

In addition to these important results, and very importantly, we have built a general methodology and computational framework for the prediction of trajectories, including stages for identifying patterns of demonstrated trajectories via clustering algorithms, identifying the features relevant to selecting different modes of behaviour via state of the art methods in ranking features, and classifying future trajectories. The important thing about this framework is that it implements a pipeline that can be automated, requiring minor human intervention, although it implements an offline process for preparing the models to be used for online trajectory prediction.

3. Conclusions, next steps and lessons learned

3.1 Conclusions

Currently, trajectory planning abilities are based on deterministic formulations of the aircraft motion problem towards making accurate predictions. Although there are sophisticated solutions that reach high levels of accuracy, all approaches are intrinsically simplifications to the actual aircraft behaviour, which delivers appropriate results for a reasonable computational cost. Although the use of the concept of Aircraft Intent together with very precise aircraft performance models such as BADA (Base of Aircraft Data) has helped to improve the prediction accuracy, the model-based approach requires a set of input data that typically are not precisely known (i.e. initial aircraft weight, pilot/FMS flight modes, …). In addition, accuracy varies depending on the intended prediction horizon (look-ahead time) and it is accurate for short horizons. Any data-driven trajectory prediction approach aims to alleviate these limitations by means of machine learning methods exploiting historical data. The potential of these methods, including Reinforcement Learning, has been explored only recently in projects such as DART [14, 16].

Starting from this point, and building on our DART experience, in this project we developed and evaluated enhanced reinforcement learning methods that are trained to imitate trajectories, treating these trajectories as rollouts of policies performed in an 4D action-state space. By doing so, we follow a supervised leaning approach, and treat historical data as data provided by "experts" that a machine learning algorithm should exploit to learn the corresponding policies comprising sequences of actions for transiting between positions in the 3D space (i.e. trajectories) through time.

Reinforcement learning techniques inherently deal with trajectories, formed as policies in an action-state space. Such methods have been used in predicting aircraft trajectories [14], as well as other types of trajectories [9,10], e.g. human and vehicle trajectories in urban spaces [12, 13], with traffic/crowd [11].

The only method dealing with aircraft trajectory prediction is the one developed in DART: This method exploits historical data on trajectories reconstructed and enhanced with aircraft intent information. Aircraft intent [15] is defined as the structured set of instructions that unambiguously specify how the aircraft is to be operated during a time interval. The AIDL is a formal language intended to express Aircraft Intent in a univocal, rigorous, and standardized manner. Language symbols are known as *instructions* and represent the minimal indivisible pieces of information that capture basic commands, guidance modes and control inputs ate the disposal of the flight deck to direct the aircraft behaviour.

Exploiting aircraft intent, the policy learned is a sequence of commands executed by the aircraft Flight Management System that produces an effect on the aircraft motion. AIDL alphabet contains 35 possible actions Exploiting aircraft intent has two major shortcomings: One is that it needs a model-based trajectory prediction method in the loop to predict the next aircraft position given a set of commands (requiring at least 500ms *at each call* to predict), while the other is that combinations of instructions capturing basic commands may not be flyable, requiring learning "constraints" on the valid commands, or approach the problem as a joint learning problem in a large state-action space. In addition to these, the approach proposed in DART discretizes the continuous state-action space (offering more "opportunities" for prediction errors), while it considers as reward the distance to the destination, which is a rather simplistic assumption.

Building on the knowledge gained from DART, we aim at building a more straightforward but novel approach in which the learning process is (a) an imitation process, where the algorithm tries to imitate "experts" planning trajectories, (b) exploiting raw trajectory data, and (c) considering unknown reward models (although arbitrarily complex) that are learned during imitation.

Our approach uses deep reinforcement learning techniques for apprenticeship learning and imitating experts: These techniques have been used in teaching robots to perform various tasks, agents to play games, and vehicles to perform autonomously, as mentioned above.

As far as we know, this is the first approach to apply deep reinforcement learning methods to imitating raw trajectories in the aviation domain, towards planning and predicting trajectories to be operated. Here, being closer to the data-driven trajectory prediction research, we applied supervised reinforcement learning for imitating experts and exploiting the models learned to predict optimal (i.e. in high-fidelity to demonstrated examples) trajectories.

Our evaluation results showed the potential and the limitations of the methods explored:

1. Deep imitation learning methods can achieve remarkable results for predicting long trajectories, also in comparison to other data-driven trajectory prediction methods.

2. They can be trained effectively, although they need many samples to explore the state-action space: However, even with a small number of demonstrated trajectories and training episodes (as our experiments in D3.1 show) they can produce highly accurate predictions.

3. These methods, as devised in this project, must be trained for any origin destination pair, requiring a considerable amount of computational resources for offline training, following the pipeline proposed.

4. Even in a singe origin destination pair with many clusters (i.e. modalities) one has to train multiple models (one per cluster). Alternatively, the method can be trained to learn a single model given all modalities: In this case the prediction accuracy is not that high as when using multiple models. Training the models is an offline process and happens once. Models can be used for online predictions very efficiently.

5. The reward (or cost) function fitted by these models does not readily reveal the features and their impact on decision making (i.e. the "true" reward/cost function used when predicting /executing trajectories). In one of our methods (GAIL), inferring such a "true" reward or cost function requires substantial effort, while in the other method (AppLearn) it should be easier, given the assumption made in this method that the reward is a linear function on specific features selected. However, under this assumption, AppLearn provides not as accurate predictions as those provided by GAIL.

3.2 Next steps

This project opens new issues for exploration for predicting trajectories using imitation learning techniques:

- Reducing the training time that the devised methods require for different origin-destination pairs: This may happen either by generalizing the state-action parameters, or by transferring knowledge learned in predicting trajectories for one pair to other pairs.

- Applying the prediction methodology to flight plans rather than to flown trajectories.

- Training different models for predicting separate trajectory segments: For instance, usually there are multiple and possibly complex landing patterns. So, we may apply the proposed prediction framework and thus, the proposed imitation learning methods to the landing segments of trajectories separately, providing more examples, training time etc. for that segment. This most probably will reveal models for landing that will be able to predict more accurately this segment of the trajectory. This of course incurs additional training cost, while it brings other complexities for combining the different trajectory segments predicted.

- Revealing the "true" reward model (e.g. measuring the cost in Euros): A way to approach this problem is to add an explanation regarding features exploited in the prediction process, which (features) are close to those features impacting the "true" reward model, as considered by airline users. This issue requires further research.

- Adding more constrains in the prediction process: E.g. availability of routes, traffic and sector congestions foreseen, etc. This will allow predictions that are much closer to the actual trajectory. We believe that this is straightforward for the proposed methods, although further exploration is necessary. Having said that, and following a data-driven approach, we need to emphasize that policies learnt should learn to satisfy constraints, given that this is the case in the examples provided. In other words, the proposed methods will learn what happens in the real world, w.r.t the constraints (e.g. in cases these are violated in reality, a data-driven model will learn to violate them, as well).

Project outcomes in terms of publications made and planned are as follows:

- Alevizos Bastas and Theocharis Kravaris and George A. Vouros, "Data Driven Aircraft Trajectory Prediction with Deep Imitation Learning", arXiv , cs.LG , 2005.07960, 2020, https://arxiv.org/abs/2005.07960 (in synergy with Alevizos Bastas' ENGAGE KTN PhD project)
- C. Spatharis, K. Blekas and George A. Vouros, "Apprenticeship learning of flight trajectories prediction with inverse reinforcement learning, submitted to SETN 2020.
- Planned submission and participation in SESAR Innovation Days 2020
- Planned Conference and Journal publication on Data Driven Aircraft Trajectory Prediction with Deep Imitation Learning.

3.3 Lessons learned

Lessons concerning management aspects

- This project, having a low overhead on managerial aspects, allowed us to concentrate on activities that advance innovative ideas bringing also experience from other projects.
- Guidance provided via assessment of project reports from mentors and Engage KTN was very helpful and useful.

Lessons concerning technical aspects

WP1:

- Data management and interlinking different data sources is always an important task requiring considerable effort: Initially we underestimated the effort, as we had implemented tools and methods for joining different sources in other projects. As new data sets are gathered, in new formats and with additional parameters, methods had to be tuned, changed or even re-implemented in some cases to address new requirements.
- Long trajectories, added a considerable effort in gathering, cleaning, managing and linking surveillance datasets to others, so as to prepare the training and testing datasets for the prediction methods.

WP2:

- Having a report on state-of-the-art methods that is updated until the end of the project was very useful: This supported tracking updates on methods and re-positioning the research so as to advance the state of the art.

- Problem formulations changed several times during the project, as we explored different alternatives also in par with the methods devised in WP3.

WP3:

- The training time required by the imitation learning method (mainly by GAIL) is considerable (in terms of 4-5 days) requiring computational power that ordinary servers or PCs cannot offer: Fortunately, we had foreseen this need and we managed to buy equipment that allowed us to run the experiments and deliver results on time. However, as said, our methods can provide more accurate results given that we provide more training time and exploration.

- Fine tuning of methods hyperparameters, in conjunction to the time needed by the experiments, is a time-consuming task that has not being performed in such a meticulous way during the project: More effort and time is needed towards this, although the tuning performed during the project shows the potential of the methods. We believe that higher accuracy on predictions can be achieved, if we tune methods hyperparameters in a meticulous way.

4. References

4.1 Project outputs

[1] D1.1 "Datasets Description", Final Edition 00.01.00, 29.11.2019,

[2] D2.1 "State of the Art Review", Final Edition 00.02.00, 29.05.2020,

[3] D3.1 "Deep Reinforcement Learning for Imitation of Trajectories", Final Edition 00.02.00, 29.05.2020

[4] Alevizos Bastas and Theocharis Kravaris and George A. Vouros, "Data Driven Aircraft Trajectory Prediction with Deep Imitation Learning", arXiv , cs.LG , 2005.07960, 2020, https://arxiv.org/abs/2005.07960 (in synergy with Alevizos Bastas' ENGAGE KTN PhD project)

[5] C. Spatharis, K. Blekas and George A. Vouros, "Apprenticeship learning of flight trajectories prediction with inverse reinforcement learning", SETN 2020.

4.2 Other

[6] P. Abbeel and A. Y Ng. 2004. Apprenticeship learning via inverse reinforcement learning. In Proceedings of the 21 st International Conference on Machine Learning, Banff, Canada, 2004.

[7] Samet Ayhan, H. S. (2016). Aircraft Trajectory Prediction Made Easy with Predictive Analytics. 22nd Intl Conf. on Knowledge Discovery and Data Mining.

[8] Delgado L., "European Route Choice Determinants", 11th USA/Europe Air Traffic Management Research and Development Seminar, Lisbon, 23 -26 June 2015, also available in https://westminsterresearch.westminster.ac.uk/download/52509e716e3e16959fe36e8c405f84d25d661a4792601ed1eb52669f3f762cb4/2452180/487_Delgado_0126150532-Final-Paper-4-30-15.pdf

[9] Bartoli F., et al “Context aware trajectory prediction”, 24th ICPR, 2018, DOI: 10.1109/ICPR.2018.8545447

[10] Pentland A., “Modeling and Prediction of Human Behavior”, Neural Computation 1999 Vol. 11, 229-242

[11] Alahi A., “Social LSTM: Human Trajectory Prediction in Crowded Spaces”, CVPR 2016.

[12] Lei Lin et al, Deep Learning-based Human-Driven Vehicle Trajectory Prediction and its Application for Platoon Control of Connected and Autonomous Vehicles”, Automated Vehicles Symposium, 2018.

[13] Truc Viet LE et al, “A reinforcement learning framework for trajectory prediction under uncertainty and budget constraint”, ECAI 2016.

[14] Boeing Research and Technology Europe filed a patent in the European Patent Office: Applicant: The Boeing Company Application title: "Method and system for autonomously operating an aircraft”, Filing date: 29th June 2017, Filing number: EP17382412.9

[15] J. Lopez-Leones, M. Vilaplana, E. Gallo, F. Navarro and C. Querejeta, “The Aircraft Intent Description Language: A key enabler for air-ground synchronization in Trajectory-Based Operations,” in Digital Avionics Systems Conference, 2007. DASC '07. IEEE/A.

[16] DART D4.5 “Final Project Results Report”, 2018, also available in http://dart-research.eu/

[17] Ho, J., & Ermon, S. (2016). Generative Adversarial Imitation Learning. Advances in neural information processing systems, (pp. 4565-4573).

[18] Yulin Liu, M. H. (2018). Predicting Aircraft Trajectories: A Deep Generative Convolutional Recurrent Neural Networks Approach. CoRR.

[19] Torabi, F., Warnell, G., & Stone, P. (2018). Generative adversarial imitation from observation. Generative adversarial imitation from observation. arXiv preprint

[20] Ribeiro, M.T, Singh S and C.Guestrin (2016). "Why Should I trust You?" Explaining the Predictions of Any Classifier, KDD 2016, DOI: http://dx.doi.org/10.1145/2939672.2939778

[21] Yang, W., Wang, K., Zuo, W. (2012). Neighborhood component feature selection for high-dimensional data. Journal of Computers, 7(1): 161-168. https://doi:10.4304/jcp.7.1.161-168